A more powerful subvector Anderson Rubin test in linear instrumental variable regression*

Patrik Guggenberger
Department of Economics
Pennsylvania State University

Frank Kleibergen
Department of Quantitative Economics
University of Amsterdam

Sophocles Mavroeidis
Department of Economics
University of Oxford

First Version: January, 2017
Revised: October 19, 2017

Abstract

We study subvector inference in the linear instrumental variables model assuming homoskedasticity but allowing for weak instruments. The subvector Anderson and Rubin (1949) test that uses chi square critical values with degrees of freedom reduced by the number of parameters not under test, proposed by Guggenberger et al (2012), controls size but is generally conservative. We propose a conditional subvector Anderson and Rubin test that uses data-dependent critical values that adapt to the strength of identification of the parameters not under test. This test has correct size and strictly higher power than the subvector Anderson and Rubin test by Guggenberger et al (2012). We provide tables with conditional critical values so that the new test is quick and easy to use.

Keywords: Asymptotic size, linear IV regression, subvector inference, weak instruments
JEL codes: C12, C26

1 Introduction

Inference in the homoskedastic linear instrumental variables (IV) regression model with possibly weak instruments has been the subject of a growing literature.1 Most of this literature has focused on the problem of inference on the full vector of slope coefficients of the endogenous regressors. Weak-instrument robust

---

*Guggenberger gratefully acknowledges the research support of the National Science Foundation via grant number SES-1326827. Mavroeidis gratefully acknowledges the research support of the European Research Council via Consolidator grant number 647152. The authors thank seminar participants at various institutions as well as Julius Koll and Jin Thed for research assistance.

inference on subvectors of slope coefficients is a harder problem, because the parameters not under test become additional nuisance parameters, and has received less attention in the literature, see e.g., Dufour and Taamouti (2005), Guggenberger et al. (2012) (henceforth GKMC), and Kleibergen (2015).

The present paper contributes to that part of the literature and focuses on the subvector Anderson and Rubin (1949) (AR) test studied by GKMC. Chernozhukov et al. (2009) showed that the full vector AR test is admissible, see also Montiel-Olea (2017). GKMC proved that the use of $\chi^2_{k-m_W}$ critical values, where $k$ is the number of instruments and $m_W$ is the number of unrestricted slope coefficients under the null hypothesis, results in a subvector AR test with asymptotic size equal to the nominal size, thus providing a power improvement over the projection approach, see Dufour and Taamouti (2005), that uses $\chi^2_k$ critical values.

This paper is motivated by the insight that the largest quantiles of the subvector AR test statistic, namely the quantiles of a $\chi^2_{k-m_W}$ distribution, occur under strong identification of the nuisance parameters. Therefore, there may be scope for improving the power of the subvector AR test by using data-dependent critical values that adapt to the strength of identification of the nuisance parameters. Indeed, we propose a new data-dependent critical value for the subvector AR test that is smaller than the $\chi^2_{k-m_W}$ critical value in GKMC. The new critical value depends monotonically on a statistic that measures the strength of identification of the nuisance parameters under the null (akin to a first-stage F statistic in a model with $m_W = 1$), and converges to the $\chi^2_{k-m_W}$ critical value when the conditioning statistic gets large. We prove that the new conditional subvector AR test has correct asymptotic size and strictly higher power than the test in GKMC, and therefore the subvector AR test in GKMC is inadmissible.

At least in the case $m_W = 1$, there is little scope for exploring alternative approaches, such as, e.g., Bonferroni, for using information about the strength of identification to improve the power of the subvector GKMC test. Specifically, in the case $m_W = 1$, we use the approach of Elliott et al. (2015) to obtain a point-optimal power bound for any test that only uses the subvector AR statistic and our measure of identification strength, and find that the power of the new conditional subvector AR test is very close to it.

Implementation of the new subvector test is trivial. The test statistic is the same as in GKMC and the critical values, as functions of a scalar conditioning statistic, are tabulated.

Our analysis relies on the insight that the subvector AR statistic is the likelihood ratio statistic for testing that the mean of a $k \times p$ Gaussian matrix with Kronecker covariance is of reduced rank. When the covariance matrix is known, this statistic corresponds to the minimum eigenvalue of a noncentral Wishart matrix. This enables us to draw on a large related statistical literature, see Muirhead (2009). A useful result from Perlman and Olkin (1980) establishes the monotonicity of the distribution of the subvector AR statistic with respect to the concentration parameter which measures the strength of identification when $m_W = 1$. The proposed conditional critical values are based on results given in Muirhead (1978) on approximations to the distribution of the eigenvalues of noncentral Wishart matrices.

In the normal linear IV model, we show that the finite-sample size of the conditional subvector AR test depends only on a $m_W$-dimensional nuisance parameter. When $m_W = 1$, it is therefore straightforward to compute the finite-sample size by simulation or numerical integration, and we prove that finite-sample size for general $m_W$ is bounded by size in the case $m_W = 1$. The conditional subvector AR test depends on eigenvalues of quadratic forms of random matrices. We combine the method of Andrews et al. (2011) that was used in GKMC with results in Andrews and Guggenberger (2015) to show that the asymptotic size of the new test can be computed from finite-sample size when errors are Gaussian and their covariance matrix is known.
Three other related papers are Rhodes Jr (1981) that studies the exact distribution of the likelihood ratio statistic for testing the validity of overidentifying restrictions in a Gaussian simultaneous equations model; and Nielsen (1999, 2001) that study conditional tests of rank in bivariate canonical correlation analysis, which is related to the present problem when \( k = 2 \) and \( m_W = 1 \). These papers do not provide results on asymptotic size or power.


The analysis in this paper relies critically on the assumption of homoskedasticity. Allowing for heteroskedasticity is difficult because the number of nuisance parameters grows with \( k \), and finite-sample distribution theory becomes intractable. When testing hypotheses on the full vector of coefficients in linear IV regression, robustness to heteroskedasticity is asymptotically costless since the heteroskedasticity-robust AR test is asymptotically equivalent to the nonrobust one under homoskedasticity, and the latter is admissible. However, in the subvector case, our paper shows that one can exploit the structure of the homoskedastic linear IV model to obtain more powerful tests, while it is not at all clear whether this is feasible under heteroskedasticity. Therefore, given the current state of the art, our results seem to indicate that there is a trade-off between efficiency and robustness to heteroskedasticity for subvector testing in the linear IV model.

The structure of the paper is as follows. Section 2 provides the finite-sample results with Gaussian errors, fixed instruments, and known covariance matrix, Section 3 gives asymptotic results, and Section 4 concludes. All proofs of the main results in the paper and tables of conditional critical values and additional numerical results are provided in the Supplemental Material (SM).

We use the following notation. For a full column rank matrix \( A \) with \( n \) rows let \( P_A = A(A'A)^{-1}A' \) and \( M_A = I_n - P_A \), where \( I_n \) denotes the \( n \times n \) identity matrix. If \( A \) has zero columns, then we set \( M_A = I_n \). The chi square distribution with \( k \) degrees of freedom and its \( 1 - \alpha \)-quantile are written as \( \chi^2_k \) and \( \chi^2_{k,1-\alpha} \), respectively. For an \( n \times n \) matrix \( A \), \( \rho(A) \) denotes the rank of \( A \) and \( \kappa_i(A), i = 1, \ldots, n \) denote the eigenvalues of \( A \) in non-increasing order. By \( \kappa_{\text{min}}(A) \) and \( \kappa_{\text{max}}(A) \) we denote the smallest and largest eigenvalue of \( A \), respectively. We write \( 0^{n \times k} \) to denote a matrix of dimensions \( n \) by \( k \) with all entries equal to zero and typically write \( 0^n \) for \( 0^{n \times 1} \).

## 2 Finite sample analysis

The model is given by the equations

\[
\begin{align*}
y &= Y\beta + W\gamma + \varepsilon \\
Y &= Z\Pi_Y + V_Y \\
W &= Z\Pi_W + V_W,
\end{align*}
\]  

(2.1)
where \( y \in \mathbb{R}^n, Y \in \mathbb{R}^{n \times m_Y}, W \in \mathbb{R}^{n \times m_W}, \) and \( Z \in \mathbb{R}^{n \times k}. \) We assume that \( k - m_W \geq 1. \) The reduced form can be written as

\[
\begin{pmatrix}
y \\
Y \\
W
\end{pmatrix} = Z \begin{pmatrix}
\Pi_Y & \Pi_W
\end{pmatrix} \begin{pmatrix}
\beta \\
\gamma
\end{pmatrix} + \begin{pmatrix}
v_y \\
V_y \\
V_W
\end{pmatrix},
\]

(2.2)

where \( v_y := V_y \beta + V_W \gamma + \varepsilon. \) By \( V_i \) we denote the \( i \)-th row of \( V \) written as a column vector and similarly for other matrices. Let \( m := m_Y + m_W. \)

Throughout this section, we make the following assumption.

**Assumption A:** 1. \( V_i := (v_{yi}, V_{yi}'_i, V_{wi}'_i)' \sim \text{i.i.d.} N \left( (0^{(m+1) \times (m+1)}, \Omega \right), i = 1, ..., n, \) where \( \Omega \in \mathbb{R}^{(m+1) \times (m+1)} \) is known and positive definite. 2. The instruments \( Z \in \mathbb{R}^{n \times k} \) are fixed and \( Z'Z \in \mathbb{R}^{k \times k} \) is positive definite.

The objective is to test the hypothesis

\[
H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0,
\]

(2.3)

using tests whose size, i.e. the highest null rejection probability (NRP) over the unrestricted nuisance parameters \( \Pi_Y, \Pi_W, \) and \( \gamma, \) equals the nominal size \( \alpha. \) In particular, weak identification and non-identification of \( \beta \) and \( \gamma \) are allowed for.

The subvector AR statistic for testing \( H_0 \) is defined as

\[
AR_n (\beta_0) := \min_{\gamma \in \mathbb{R}^{m_W}} \frac{(\nabla_0 - W\gamma)'P_Z(\nabla_0 - W\gamma)}{(1, -\gamma')(\Omega (\beta_0) (1, -\gamma')'),
\]

(2.4)

where

\[
\Omega (\beta_0) := \begin{pmatrix}
1 & 0^{1 \times m_W} \\
-\beta_0 & 0^{m_Y \times m_W} \\
0^{m_W \times 1} & I_{m_W}
\end{pmatrix},
\]

(2.5)

and

\[
\nabla_0 := y - Y\beta_0.
\]

(2.6)

Denote by \( \hat{\kappa}_i \) for \( i = 1, ..., p := 1 + m_W \) the roots of the following characteristic polynomial in \( \kappa \)

\[
| \kappa \Omega (\beta_0) - (\nabla_0, W)'P_Z(\nabla_0, W) | = 0,
\]

(2.7)

ordered non-increasingly. Then,

\[
AR_n (\beta_0) = \hat{\kappa}_p,
\]

(2.8)

that is, \( AR_n (\beta_0) \) equals the smallest characteristic root, see, e.g. (Schmidt, 1976, chapter 4.8). The subvector AR test in GKMC rejects \( H_0 \) at significance level \( \alpha \) if \( AR_n (\beta_0) > \chi^2_{k - m_W, 1 - \alpha}, \) while the AR test based on projection rejects if \( AR_n (\beta_0) > \chi^2_{k, 1 - \alpha}. \)

Under Assumption A, the subvector AR statistic equals the minimum eigenvalue of a noncentral Wishart matrix. More precisely, we show in the SM (Subsection S.1.1) that the roots \( \hat{\kappa}_i \) of (2.7) for \( i = 1, ..., p \), satisfy

\[
0 = |\hat{\kappa}_i I_p - \Xi |,
\]

(2.9)

where \( \Xi \sim N (\mathcal{M}, I_{kp}) \) for some nonrandom \( \mathcal{M} \in \mathbb{R}^{k \times p} \) (defined in (S-15) in the SM). Furthermore, under the
null hypothesis $H_0$, $\mathcal{M} = (0^k, \Theta_W)$ for some $\Theta_W \in \mathbb{R}^{k \times m_W}$ (defined in (S–17) in the SM) and thus $\rho(\mathcal{M}) \leq m_W$. Therefore, $\Xi' \Xi \sim \mathcal{W}_p(k, I_p, \mathcal{M}' \mathcal{M})$, where the latter denotes a non-central Wishart distribution with $k$ degrees of freedom, covariance matrix $I_p$, and noncentrality matrix

$$\mathcal{M}' \mathcal{M} = \begin{pmatrix} 0 & 0^{1 \times m_W} \\ 0^{m_W \times 1} & \Theta_W' \Theta_W \end{pmatrix}. \quad (2.10)$$

The joint distribution of the eigenvalues of a noncentral Wishart matrix only depends on the eigenvalues of the noncentrality matrix $\mathcal{M}' \mathcal{M}$ (see e.g. Muirhead, 2009). Hence, the distribution of $(\hat{\kappa}_1, \ldots, \hat{\kappa}_p)$ under the null only depends on the eigenvalues of $\Theta_W' \Theta_W$, which we denote by

$$\kappa_i := \kappa_i (\Theta_W' \Theta_W), \ i = 1, \ldots, m_W. \quad (2.11)$$

We can think of $\Theta_W' \Theta_W$ as the concentration matrix for the endogenous regressors $W$, see e.g. Stock et al. (2002). In the case when $m_W = 1$, $\Theta_W' \Theta_W$ is a scalar, and corresponds to the well-known concentration parameter (see e.g. Staiger and Stock (1997)) that measures the strength of the identification of the parameter vector $\gamma$ not under test.

### 2.1 Motivation for conditional subvector AR test: Case $m_W = 1$

The above established that when $m_W = 1$ the distribution of $AR_n (\beta_0)$ under $H_0$ depends only on the single nuisance parameter $\kappa_1$. The following result gives a useful monotonicity property of this distribution.

**Theorem 1** Suppose that Assumption A holds and $m_W = 1$. Then, under the null hypothesis $H_0 : \beta = \beta_0$, the distribution function of the subvector AR statistic in (2.4) is monotonically decreasing in the parameter $\kappa_1$, defined in (2.11), and converges to $\chi^2_{k-1}$ as $\kappa_1 \to \infty$.

This result follows from (Perlman and Olkin, 1980, Theorem 3.5), who established that the eigenvalues of a $k \times p$ noncentral Wishart matrix are stochastically increasing in the nonzero eigenvalue of the noncentrality matrix when the noncentrality matrix is of rank 1.

Theorem 1 shows that the subvector AR test in GKMC is conservative for all $\kappa_1 < \infty$, because its NRP $Pr_{\kappa_1} (AR_n (\beta_0) > \chi^2_{k-1,1-\alpha})$ is monotonically increasing in $\kappa_1$ and the worst case occurs at $\kappa_1 = \infty$. Hence, it seems possible to improve the power of the subvector AR test by reducing the $\chi^2_{k-1}$ critical value based on information about the value of $\kappa_1$.

If $\kappa_1$ were known, which it is not, one would set the critical value equal to the $1 - \alpha$ quantile of the exact distribution of $AR_n (\beta_0)$ and obtain a similar test with higher power than the subvector AR test in GKMC. Alternatively, if there was a one-dimensional minimal sufficient statistic for $\kappa_1$ under $H_0$, one could obtain a similar test by conditioning on it. Unfortunately, we are not aware of such a statistic. However, an approximation to the density of eigenvalues of noncentral Wishart matrices by Leach (1969), specialized to this case, implies that the largest eigenvalue $\hat{\kappa}_1$ is approximately sufficient for $\kappa_1$ when $\kappa_1$ is “large” and $\kappa_2 = 0$. Based on this approximation, (Muirhead, 1978, Section 6) provides an approximate, nuisance parameter free, conditional density of the smallest eigenvalue $\hat{\kappa}_2$ given the largest one $\hat{\kappa}_1$. This approximate density (with respect to Lebesgue measure) of $\hat{\kappa}_1$ can be written as

$$f^{\hat{\kappa}_2|\hat{\kappa}_1}_{\hat{\kappa}_1} (x_2 | \hat{\kappa}_1) = f_{\chi^2_{k-1}} (x_2) (\hat{\kappa}_1 - x_2)^{1/2} g (\hat{\kappa}_1), \quad x_2 \in [0, \hat{\kappa}_1], \quad (2.12)$$
the limiting result is given in Section S.1.2 in the SM.

The polynomial in (2.7) is comparable to the first-stage F statistic in the case a close to similar test, except for small values of.

Values have been rounded up to one decimal.

Figure 1: Conditional quantile function. The solid line plots the $1 - \alpha$ quantile of the distribution with density (2.12), for $\alpha = 5\%$. The dotted straight blue line gives the corresponding quantile of $\chi_1^{2}$.

where $f_{\chi_1^{2}}(\cdot)$ is the density of a $\chi_1^{2}$ and $g(\hat{\kappa}_1)$ is a function that does not depend on any unknown parameters, see (S–26) in the SM.

Because (2.12) is analytically available, the quantiles of the distribution whose density is given in (2.12) can be computed easily using numerical integration for fixed values of $\hat{\kappa}_1$. Figure 1 plots the $1 - \alpha$ quantile of that distribution as a function of $\hat{\kappa}_1$ for $\alpha = 5\%$ and $k = 2, 5, 10, \text{and } 20$. It is evident that this conditional quantile function is strictly increasing in $\hat{\kappa}_1$ and asymptotes to $\chi_1^{2}$. We propose to use the above conditional quantile function to obtain conditional critical values for the subvector AR statistic.

In practice, to make implementation of the test straightforward for empirical researchers, the conditional critical value function will be tabulated for different $k - 1$ and $\alpha$ over a grid of points $\hat{\kappa}_{1,j}$, $j = 1, \ldots, J$, say, and conditional critical values for any given $\hat{\kappa}_1$ will be obtained by linear interpolation. Specifically, let $q_{1-\alpha,j}(k-1)$ denote the $1 - \alpha$ quantile of the distribution whose density is given by (2.12) with $\hat{\kappa}_1$ replaced by $\hat{\kappa}_{1,j}$. The end point of the grid $\hat{\kappa}_{1,j}$ should be chosen high enough so that $q_{1-\alpha,j}(k-1) \approx \chi_1^{2}$. For any realization of $\hat{\kappa}_1 \leq \hat{\kappa}_{1,j}$, find $j$ such that $\hat{\kappa}_1 \in [\hat{\kappa}_{1,j-1}, \hat{\kappa}_{1,j}]$ with $\hat{\kappa}_{1,0} = 0$ and $q_{1-\alpha,0}(k-1) = 0$, and let

$$c_{1-\alpha}(\hat{\kappa}_1, k - 1) := \frac{\hat{\kappa}_{1,j} - \hat{\kappa}_1}{\hat{\kappa}_{1,j} - \hat{\kappa}_{1,j-1}} q_{1-\alpha,j-1}(k-1) + \frac{\hat{\kappa}_1 - \hat{\kappa}_{1,j-1}}{\hat{\kappa}_{1,j} - \hat{\kappa}_{1,j-1}} q_{1-\alpha,j}(k-1).$$  \hspace{1cm} (2.13)

Table 1 gives conditional critical values at significance level 5\% for a fine grid for the conditioning statistic $\hat{\kappa}_1$ for the case $k - 1 = 4$. To mitigate any slight over-rejection induced by interpolation, the reported critical values have been rounded up to one decimal.

We will see that by using $c_{1-\alpha}(\hat{\kappa}_1, k - 1)$ as a critical value for the subvector AR test, one obtains a close to similar test, except for small values of $\kappa_1$. Note that $\hat{\kappa}_1$, the largest root of the characteristic polynomial in (2.7) is comparable to the first-stage F statistic in the case $m_W = 1$ for the hypothesis that

\hspace{1cm} \footnote{The monotonicity statement is made based on numerical integration without an analytical proof. An analytical proof of the limiting result is given in Section S.1.2 in the SM.}

\footnote{For general $m_W$, discussed in the next subsection, the role of $k - 1$ is played by $k - m_W$.}

\footnote{When $\hat{\kappa}_1 > \hat{\kappa}_{1,j}$, we can define $c_{1-\alpha}(\hat{\kappa}_1, k - 1)$ using nonlinear interpolation between $\hat{\kappa}_{1,j}$ and $\infty$, i.e., $c_{1-\alpha}(\hat{\kappa}_1, k - 1) := (1 - F(\hat{\kappa}_1 - \hat{\kappa}_{1,j})) q_{1-\alpha,j}(k-1) + F(\hat{\kappa}_1 - \hat{\kappa}_{1,j}) \chi_1^{2}$, where $F$ is some distribution function.}
To provide an analytical expression of the NRP, we need to do that numerically. This can be done easily because the values of the conditioning variable \( \hat{\kappa}_1 \) are the main purpose of (2.12) was to give us a simple analytical expression to generate data-dependent critical values for the approximation (2.12) to the true distribution. As long as the conditional critical values are identified, the nonnegative scalar \( \hat{\kappa}_1 \) which is a measure of the strength of identification of the unrestricted coefficients.

### 2.2 Definition of the conditional subvector AR test for general \( m_W \)

We will now define the conditional subvector AR test for the general case when \( m_W \geq 1 \). The conditional subvector AR test rejects \( H_0 \) at nominal size \( \alpha \) if

\[
AR_\alpha(\beta_0) > c_{1-\alpha}(\hat{\kappa}_1, k - m_W),
\]

where \( c_{1-\alpha}(\cdot, \cdot) \) has been defined in (2.13) for any argument consisting of a vector with first component in \( \mathbb{R}_+ \cup \{\infty\} \) and second component in \( \mathbb{N} \). Tables of critical values for significance levels \( \alpha = 10\%, 5\%, \) and \( 1\% \), and degrees of freedom \( k - m_W = 1 \) to 20 are provided in Section S.3 of the SM. Since \( AR_\alpha(\beta_0) = \hat{\kappa}_p \), the associated test function can be written as

\[
\varphi_c(\hat{\kappa}) := \mathbf{1}[\hat{\kappa}_p > c_{1-\alpha}(\hat{\kappa}_1, k - m_W)],
\]

where \( \mathbf{1}[\cdot] \) is the indicator function, \( \hat{\kappa} := (\hat{\kappa}_1, \hat{\kappa}_p) \) and the subscript \( c \) abbreviates “conditional”.

The subvector AR test in GKMC that uses \( \chi^2_{k-m_W} \) critical value has test function

\[
\varphi_{GKMC}(\hat{\kappa}) := \mathbf{1}[\hat{\kappa}_p > c_{1-\alpha}(\infty, k - m_W)].
\]

Since \( c_{1-\alpha}(x, \cdot) < c_{1-\alpha}(\infty, \cdot) \) for all \( x < \infty \), it follows that \( E[\varphi_c(\hat{\kappa})] > E[\varphi_{GKMC}(\hat{\kappa})] \), i.e., the conditional subvector AR test \( \varphi_c \) has strictly higher power than the (unconditional) subvector AR test \( \varphi_{GKMC} \) in GKMC.

### 2.3 Finite sample size of \( \varphi_c \) when \( m_W = 1 \)

As long as the conditional critical values \( c_{1-\alpha}(\hat{\kappa}_1, k - m_W) \) guarantee size control for the new test \( \varphi_c \), the actual quality of the approximation (2.12) to the true conditional density is not of major concern to us, and the main purpose of (2.12) was to give us a simple analytical expression to generate data-dependent critical values.

We next compute the size of the conditional subvector AR test, and because we don’t have available an analytical expression of the NRP, we need to do that numerically. This can be done easily because the
nuisance parameter $\kappa_1$ is one-dimensional, and the density of the data is analytically available, so the NRP of the test can be estimated accurately by Monte Carlo simulation or numerical integration. Using (low-dimensional) simulations to calculate the (asymptotic) size of a testing procedure has been used in several recent papers, see e.g. Elliott et al. (2015).

Figure 2 plots the NRPs of both $\varphi_c$ and the subvector AR test $\varphi_{GKMC}$ of GKMC in (2.16) at $\alpha = 5\%$ as a function of $\kappa_1$ for $k = 5$ and $m_W = 1$. The conditional test $\varphi_c$ is evaluated using the critical values reported in Table 1 with interpolation.

We notice that the size of the conditional subvector AR test $\varphi_c$ is controlled, because the NRPs never exceed the nominal size no matter the value of $\kappa_1$. The NRPs of the subvector AR test $\varphi_{GKMC}$ are monotonically increasing in $\kappa_1$ in accordance with Theorem 1. Therefore the proposed conditional test $\varphi_c$ strictly dominates the unconditional test $\varphi_{GKMC}$. The results for other significance levels and other values of $k$ are the same, and they are reported in Table S.21 of the SM. We summarize this finding in the following theorem.

**Theorem 2** Under Assumption A, the finite-sample size of the conditional subvector AR test $\varphi_c$ defined in (2.15) is equal to its nominal size $\alpha$.

**Comment.** To reiterate, the proof of Theorem 2 for given $k - m_W$ and nominal size $\alpha$ is a combination of an analytical step that shows that the null rejection probability of the new test depends on only a scalar parameter and of a numerical step where it is shown by numerical integration and Monte Carlo simulation that none of the NRPs exceeds the nominal size. We performed these simulations for $k - m_W = 1, \ldots, 20$ and $\alpha = 10\%, 5\%$, and $1\%$ using 1 million Monte Carlo replications, and in no case did we find size distortion.

---

5E.g. if $\hat{\kappa}_1 = 2.4$ which is an element of $[2.3, 2.5]$, then from Table 1 the critical value employed would be 2.2. To produce Figure 2 we use a grid of 42 points for $\kappa_1$, evenly spaced in log-scale between 0 and 100. In this figure, the NRPs were computed by numerical integration using the Quadpack in Ox, see Doornik (2001). The densities were evaluated using the algorithm of Koev and Edelman (2006) for the computation of hypergeometric functions of two matrix arguments. The NRPs are essentially the same when estimated by Monte Carlo integration with 1 million replications, see Section S.2 in the SM.
## 2.4 Power analysis when $m_W = 1$

One main advantage of the conditional subvector AR test (2.14) is its computational simplicity. For general $m_W$ there are other approaches one might consider based on the information in the eigenvalues $(\hat{\kappa}_1, \ldots, \hat{\kappa}_{m_W})$ that, at the expense of potentially much higher computational cost, might yield higher power than the conditional subvector AR test. For example, one could apply the critical value function approach of Moreira et al. (2016) to derive conditional critical values. One could condition on the largest $m_W$ eigenvalues rather than just the largest one.

To assess the scope for power improvements over the subvector AR test in GKMC, we consider the case $m_W = 1$ and compute power bounds of all tests that depend on the statistic $(\hat{\kappa}_1, \hat{\kappa}_2)$. These are point-optimal bounds based on the least favorable distribution for the nuisance parameter $\kappa_1$ under the null that $\kappa_2 = 0$, see the SM S.2.3 for details. We consider both the approximately least favorable distribution (ALFD) method of Elliott et al. (2015) and the one-point least favorable distribution of (Andrews et al., 2008, section 4.2), but report here only the ALFD bound for brevity and because it is very similar to the Andrews et al. (2008) upper bound. The results based on the Andrews et al. (2008) method are discussed in Section S.4.2 of the SM.

Recall from (2.11) that under $H_0 : \beta = \beta_0$ in (2.3), the joint distribution of $(\hat{\kappa}_1, \ldots, \hat{\kappa}_p)$ only depends on the vector of eigenvalues $(\kappa_1, \ldots, \kappa_{m_W})$ of $\Theta_W$, where $\Theta_W \in \mathbb{R}^{k \times m_W}$ appears in the noncentrality matrix $\mathcal{M} = (0^k, \Theta_W)$ of $\Xi \sim N(\mathcal{M}, I_p)$. It follows easily from (S–17) in the SM that if $\Pi_W$ ranges through all matrices in $\mathbb{R}^{k \times m_W}$, then $(\kappa_1, \ldots, \kappa_{m_W})'$ ranges through all vectors in $[0, \infty)^{m_W}$.

Define $A := E(Z'(y - Y_0, W)) \in \mathbb{R}^{k \times p}$ and consider the null hypothesis

$$H'_0 : \rho(A) \leq m_W \text{ versus } H''_1 : \rho(A) = p. \quad (2.17)$$

Clearly, whenever $H_0$ holds $H'_0$ holds too, but the reverse is not true; by (S–18) in the SM, $H'_0$ holds iff $\Pi_W$ is rank deficient or $\Pi_Y(\beta - \beta_0) \in \text{span}(\Pi_W)$. It is shown in the SM (Case 2 in Subsection S.1.1) that under $H'_0$ the joint distribution of $(\hat{\kappa}_1, \ldots, \hat{\kappa}_p)$ is the same as the one of the vector of eigenvalues of a Wishart matrix $W_p(k, I_p, \mathcal{M}'\mathcal{M})$ with rank deficient noncentrality matrix and therefore depends only on the vector of the largest $m_W$ eigenvalues $(\kappa_1, \ldots, \kappa_{m_W})'$ in $\mathbb{R}^{m_W}$ of $\mathcal{M}'\mathcal{M}$. The important implication of that result is that any test $\varphi(\kappa_1, \ldots, \kappa_{m_W}) \in [0, 1]$ for some measurable function $\varphi$ that has size bounded by $\alpha$ under $H_0$ also has size (in the parameters $(\beta, \gamma, \Pi_Y, \Pi_W)$) bounded by $\alpha$ under $H'_0$. In particular, no test $\varphi(\hat{\kappa}_1, \ldots, \hat{\kappa}_p)$ that controls size under $H_0$ has power exceeding size under alternatives $H'_0 \backslash H_0$.

Now assume $m_W = 1$. We compute the power of the conditional and unconditional subvector tests $\varphi_c$ and $\varphi_{GKMC}$ at the 5% level for $k = 5$ and the associated power bound for a grid of values of the parameters $\kappa_1 \geq \kappa_2 > 0$ under the alternative, see Section S.2.3 in the SM for details. The power curves are computed using 100,000 Monte Carlo replications without importance sampling (results for other $k$ are similar and given in the SM). The left panel of Figure 3 plots the difference between the power function of the conditional test $\varphi_c$ and the power bound across all alternatives. Except at alternatives very close to the null, and when $\kappa_1$ is very close to $\kappa_2$ (so the nuisance parameter is weakly identified), the power of the conditional subvector test $\varphi_c$ is essentially on the power bound. The fact that the power of $\varphi_c$ for small $\kappa_1$ is somewhat below the power bound can be explained by the fact that the test is not exactly similar, so its rejection probability can fall below $\alpha$ for some alternatives. The right panel of Figure 3 plots the power curves for alternatives with $\kappa_1 = \kappa_2$, which seem to be the least favorable to the conditional test. The power of the conditional test is mostly on the power bound, while the subvector test $\varphi_{GKMC}$ is well below the bound. Two-dimensional
Power of $\phi_c$ minus power bound

\[ \kappa_2 - \kappa_1 - 25 \]

\[ 0.02, 0.01, 0 \]

\[ 75, 50, 25, 10, 0, 20, 30 \]

Power curves when $\kappa_1 = \kappa_2$

\[ 0.0, 0.5, 1.0 \]

\[ 0, 5, 10, 15, 20, 25, 30 \]

Figure 3: Power of conditional (2.15) and GKMC (2.16) subvector AR tests, $\varphi_c$ and $\varphi_{GKMC}$, and point optimal power envelope computed using the ALFD method of Elliott et al. (2015). The number of instruments is $k = 5$ and the number of nuisance parameters is $m_W = 1$. The left panel plots the power of $\varphi_c$ minus the power bound across all alternatives. The right panel plots the power curves for both tests and the power bound when $\kappa_1 = \kappa_2$. Plots for other values of $\kappa_1 - \kappa_2$ are provided in the SM. As $\kappa_1 - \kappa_2$ gets larger, the power of $\varphi_{GKMC}$ gets closer to the power envelope, as expected.

2.5 Size of $\varphi_c$ when $m_W > 1$ - inadmissibility of $\varphi_{GKMC}$

We cannot extend the monotonicity result of Theorem 1 to the general case $m_W > 1$. This is because the distribution of the subvector AR statistic depends on all the $m_W$ eigenvalues of $M'M$ in (2.10), and the method of the proof of Theorem 1 only works for the case that $\rho(M'M) = 1$.

However, the following result suffices to establish correct finite-sample size of the proposed conditional subvector AR test (2.15) and the inadmissibility of the subvector test $\varphi_{GKMC}$ in (2.16) in the general case.

**Theorem 3** Suppose that Assumption A holds with $m_W > 1$. Denote by $\tilde{\Xi}_{11} \in \mathbb{R}^{k-m_W+1 \times 2}$ the upper left submatrix of $\tilde{\Xi} := \Xi O \in \mathbb{R}^{k \times p}$, where $\Xi$ and the random orthogonal matrix $O \in \mathbb{R}^{p \times p}$ are defined below (2.9) and in (S-4) of the SM, respectively. Then, under the null hypothesis $H_0 : \beta = \beta_0$

\[ \tilde{\Xi}_{11} \tilde{\Xi}_{11} | O \sim W_2 \left(k - m_W + 1, I_2, \hat{\mathcal{M}}_{11}, \tilde{\mathcal{M}}_{11}\right), \]

where $\hat{\mathcal{M}}_{11}$ is defined in (S-7) in the SM and satisfies $\rho(\hat{\mathcal{M}}_{11}) \leq 1$.

Note that

\[ AR_n(\beta_0) = \kappa_{\min}(\Xi'\Xi) = \kappa_{\min}(\tilde{\Xi}'\tilde{\Xi}) \leq \kappa_{\min}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) \leq \kappa_{\max}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) \leq \kappa_{\max}(\Xi'\Xi) = \kappa_{\max}(\Xi'\Xi), \]

(2.18)

where the first and third inequalities hold by the inclusion principle, see (Lütkepohl, 1996, p. 73) and the second and last equalities hold because $O$ is orthogonal. Therefore,

\[ P(AR_n(\beta_0) > c_{1-\alpha}(\kappa_{\max}(\Xi'\Xi), k - m_W)) \leq P(\kappa_{\min}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) > c_{1-\alpha}(\kappa_{\max}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}), k - m_W)) \leq \alpha, \]

(2.19)

\[ ^6\text{See (Perlman and Olkin, 1980, p. 1337) for some more discussion of the difficulties involved in extending the result to the general case.} \]
where the first inequality follows from (2.18). The second inequality follows from Theorem 2 for the case $m_W = 1$ and from Theorem 3 by conditioning on $O$, where the role of $k$ is now played by $k - m_W + 1$. Hence, the conditional subvector AR test has correct size for any $m_W$. Because $c_1 - \alpha(k_{\text{max}}(\Xi^T \Xi), k - m_W) < \chi^2_{k - m_W, 1 - \alpha}$, it follows that the subvector AR test $\varphi_{GKMC}$ given in (2.16) is inadmissible.

2.6 Refinement

Figure 2 shows that the NRPs of test $\varphi_c$ for nominal size 5% is considerably below 5% for small values of $\kappa_1$, which causes a loss of power for some alternatives that are close to $H_0$, see Figure 3. However, we can reduce the under-rejection by adjusting the conditional critical values to bring the test closer to similarity.\footnote{We thank Ulrich Müller for this suggestion.}

For the case $k = 5$, $m_W = 1$, and $\alpha = 5\%$, let $\varphi_{adj}$ be the test that uses the critical values in Table 1 where the smallest 8 critical values are divided by 5 (e.g., the critical value for $\kappa_1 = 2.5$ becomes 0.46). Figure 4 shows that $\varphi_{adj}$ still has size 5%, that it is much closer to similarity than $\varphi_c$, and does not suffer from any loss of power relative to the power bound near $H_0$. This approach can be applied to all other values of $\alpha$ and $k$, but needs to be adjusted for each case.

3 Asymptotics

In this section, Assumption A is replaced by

**Assumption B:** The random vectors $(\varepsilon_i, Z_i, V_{Y,i}, V_{W,i})$ for $i = 1, \ldots, n$ in (2.1) are i.i.d. with distribution $F$.

Therefore, the instruments are random, the reduced form errors are not necessarily normally distributed, and the matrix $\Omega = E_{F} V_i V_i^T$ is unknown. We define the parameter space $\mathcal{F}$ for $(\gamma, \Pi_W, \Pi_Y, F)$ under the null hypothesis $H_0 : \beta = \beta_0$ exactly as in GKMC.\footnote{Regarding the notation $(\gamma, \Pi_W, \Pi_Y, F)$ and elsewhere, note that we allow as components of a vector column vectors, matrices (of different dimensions), and distributions.} Namely, for $U_i = (\varepsilon_i, V_{W,i})'$ let
\[ F = \{ (\gamma, \Pi_W, \Pi_Y, F) : \gamma \in \mathbb{R}^{nW}, \Pi_W \in \mathbb{R}^{k \times mW}, \Pi_Y \in \mathbb{R}^{k \times mY}, \]
\[ E_F(||T_i||^{2+\delta}) \leq B, \quad \text{for } T_i \in \{Z_i\varepsilon_i, vec(Z_iV_{W,i}), V_{W,i}\varepsilon_i, \varepsilon_i, V_{W,i}, Z_i\}, \]
\[ E_F(Z_iV_i^2) = 0^{k \times (m+1)}, \quad E_F(vec(Z_iU_i')(vec(Z_iU_i'))) = (E_F(U_iU_i') \otimes E_F(Z_iZ_i')), \]
\[ \kappa_{\min}(A) \geq \delta \text{ for } A \in \{E_F(Z_iZ_i'), E_F(U_iU_i')\} \]  

(3.1)

for some \( \delta > 0 \) and \( B < \infty \), where "\( \otimes \)" denotes the Kronecker product of two matrices and \( vec(\cdot) \) the column vectorization of a matrix. Note that the factorization of the covariance matrix into a Kronecker product in line three of (3.1) is a slightly weaker assumption than conditional homoskedasticity.

Rather than controlling the finite-sample size the objective is to demonstrate that the new conditional subvector AR test has \textit{asymptotic size}, that is the limit of the finite-sample size with respect to \( F \), equal to the nominal size.

We next define the test statistic and the critical value for the case here where \( \Omega \) is unknown. With some abuse of notation (by using the same symbol for another object than above), the subvector AR statistic \( AR_n(\beta_0) \) is defined as the smallest root \( \hat{\kappa}_{pn} \) of the roots \( \hat{\kappa}_{in}, i = 1, \ldots, p \) (ordered nonincreasingly) of the characteristic polynomial

\[
|\hat{\kappa}I_p - \hat{U}_n (\overline{Y}_0, W)' P_Z (\overline{Y}_0, W) \hat{U}_n | = 0, \tag{3.2}
\]

where

\[
\hat{U}_n := ((n - k)^{-1} (\overline{Y}_0, W)' M_Z (\overline{Y}_0, W))^{-1/2} \tag{3.3}
\]

and \( \hat{U}_n^{-2} \) is a consistent estimator (under certain drifting sequences from the parameter space \( F \)) for \( \Omega (\beta_0) \) in (2.5), see Lemma 1 in the SM for details. The conditional subvector AR test rejects \( H_0 \) at nominal size \( \alpha \) if

\[
AR_n(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W), \tag{3.4}
\]

where \( c_{1-\alpha}(\cdot, \cdot) \) has been defined in (2.13) and \( \hat{\kappa}_{1n} \) is the largest root of (3.2).

\textbf{Theorem 4} Under Assumption B, the conditional subvector AR test in (3.4) implemented at nominal size \( \alpha \in (0, 1) \) has asymptotic size equal to \( \alpha \) for the parameter space \( F \) defined in (3.1).

\textbf{Comments.} 1. The proof of Theorem 4 is given in Section S.1.3 in the SM. It relies on showing that the limiting NRP is smaller or equal to \( \alpha \) along all relevant drifting sequences of parameters from \( F \). This is done by showing that the limiting NRPs equal finite-sample NRPs under Assumption A. Therefore the same comment applies to Theorem 4 as the comment below Theorem 2. The analysis is substantially more complicated here than in GKMC, in part because the critical values are also random.

2. Theorem 4 remains true if the conditional critical value \( c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) \) of the subvector AR test is replaced by any other critical value, \( \overline{c}_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) \) say, where \( \overline{c}_{1-\alpha}(\cdot, k - m_W) \) is a continuous non-decreasing function such that the corresponding test under Assumption A has finite-sample size equal to \( \alpha \). In particular, besides the critical values obtained from Table 1 by interpolation also the critical values suggested in Section 2.6 could be used.
4 Conclusion

We show that the subvector AR test of GKMC is inadmissible by developing a new conditional subvector AR test that has correct size and uses data-dependent critical values that are always smaller than the $\chi_{k-m_W}^2$ critical values in GKMC. The critical values are increasing in a conditioning statistic that relates to the strength of identification of the parameters not under test. Our proposed test has considerably higher power under weak identification than the GKMC procedure.

References


